APPLICATION OF ENSEMBLE DETECTION AND ANALYSIS TO MODELING UNCERTAINTY IN NON STATIONARY PROCESSES

Paul Racette, DSc
NASA Goddard Space Flight Center

ABSTRACT

Characterization of non stationary and nonlinear processes is a challenge in many engineering and scientific disciplines. Climate change modeling and projection, retrieving information from Doppler measurements of hydrometeors, and modeling calibration architectures and algorithms in microwave radiometers are example applications that can benefit from improvements in the modeling and analysis of non stationary processes.

Analyses of measured signals have traditionally been limited to a single measurement series. Ensemble Detection is a technique whereby mixing calibrated noise produces an ensemble measurement set. The collection of ensemble data sets enables new methods for analyzing random signals and offers powerful new approaches to studying and analyzing non stationary processes. Derived information contained in the dynamic stochastic moments of a process will enable many novel applications.

Index Terms— Radiometer calibration, measurement uncertainty, non stationary processes, ensemble detection and analysis, empirical mode decomposition, observation theory

1. INTRODUCTION

All elements of Nature exhibit characteristics that change when considered over sufficiently large temporal or spatial scales. Considerable effort and resources are applied to 'designing around' non stationary and nonlinear effects. Nonlinear effects of observation methodology pose the most perplexing aspects for modeling non stationary processes. The statistics one obtains depends upon the sampling of the process. Consider, for example, the difficulties associated with averaging high-resolution radar data to match coarser-resolution passive microwave measurements for combined active/passive retrievals of soil moisture. The lack of well-developed techniques for modeling changing statistical moments in our observations has stymied the application of stochastic process theory in science and engineering. These limitations were encountered when modeling temporal effects of calibration frequency on the performance of a radiometer with non stationary receiver fluctuations. The mathematics for modeling and analyzing radiometer data yields Ensemble Detection and Analysis (EDA), a novel form of noise assisted data analysis for measuring and modeling non stationary processes.

2. MEASUREMENTS USING CALIBRATED NOISE

Calibration provides the condition by which assignment of value can be made and the means of discriminating a signal from background noise. Calibration using standard references also provides the means by which we can compare measurements across space and time. These properties make calibration applicable to detecting non stationary processes and analyzing how their statistical moments change with time, location, interval, etc…

Racette and Lang (2005) [1] describe how measurement uncertainty can be used as a figure of merit in comparing radiometer calibration architectures and algorithms. The method uses stochastic process theory to analytically evaluate the uncertainty in estimating a measurand value, T_A, using other references for calibration. The uncertainty depends on reference and measurand values, calibration frequency, estimation algorithm and underlying non stationary properties of the receiver fluctuations. The mathematical formulation allows direct comparison of model calculations to data analysis as shown in Racette (2005) [2].

A generic radiometer design is shown in Figure 1. The radiometer sequentially samples the calibration references and the measurand. Receiver gain fluctuations appear in each of the reference measurement sequences; each measurement is a realization of the receiver gain, thus the sequences comprise an ensemble set of the gain fluctuations. Following the development in [1] for the case of stationary receiver fluctuations, the calculation of uncertainty is made by first forming a set of characteristic equations

$$
\begin{aligned}
T_A &= m(t_A)v_A(t_A) + b(t_A) + \varepsilon_A(t_A, t_A) \\
T_1 &= m(t_A)v_1(t_1) + b(t_A) + \varepsilon_1(t_A, t_1) \\
T_2 &= m(t_A)v_2(t_2) + b(t_A) + \varepsilon_2(t_A, t_2) \\
&\vdots \\
T_n &= m(t_A)v_n(t_n) + b(t_A) + \varepsilon_n(t_A, t_n)
\end{aligned}
$$

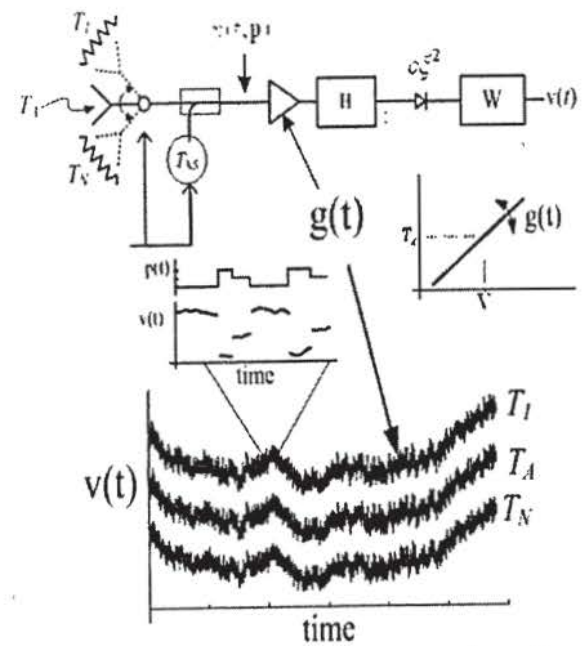


Figure 1: Diagram illustrating the output of a radiometer with periodic calibration. Gain fluctuations in the receiver appear in each of the noise reference measurements. The reference measurements comprise an ensemble set of receiver gain.

where $\varepsilon(t_A, t_i)$ is the deviation of sample pair $\{v_i, T_i\}$ from the mean receiver state at t_A, i.e. $m(t_A)$ and $b(t_A)$ are the *mean* slope and offset of the receiver response at t_A. The times, t_1, t_2, ... t_n, are the times at which calibration measurements are made. This set of characteristic equations are used in a calibration algorithm **f** to form an estimate of the receiver state at t_A and an estimate of T_A,

$$\hat{T}_A = \hat{m}(t_A)v_A(t_A) + \hat{b}(t_A).$$

Figure 2 illustrates an algorithm using two calibration reference pair measurements projected in time to form an estimate of $\hat{T}_A$ at t_A. The estimate $\hat{T}_A$ is calculated from set of radiometer sample voltages, **v**, and calibrated reference temperature, **T**, over an interval L_{obs}. The variance of the estimator is calculated by applying the calibration algorithm over an interval L. Typically, algorithms are designed to minimize the affect of receiver fluctuations on the estimate. However, information about the stochastic properties of the fluctuations can be derived by varying the temporal structure of the algorithm. To the extent that the receiver and references are stable, the statistical moments of $\hat{T}_A$ do not depend on the temporal spacing of the reference samples. The stationary statistics thus form form boundary conditions against which deviations from the stationary assumption can be detected.

3. MODELING UNCERTAINTY OF NON STATIONARY PROCESSES

The uncertainty associated with the estimator $\hat{T}_A$ is given by

$$\sigma^2_{\hat{T}_A} = E\left\{\left(\hat{T}_A - T_A\right)^2\right\}$$

where E{} is the ensemble expectation operator. Evaluation of the estimator uncertainty is developed in [1] for stationary fluctuations in g(t). The development for non stationary fluctuations follows along the same lines and converges to the results present in [1] for the stationary case. The uncertainty can be evaluated for algorithms in which **f** is analytic. Expanding $\hat{T}_A$ into a multivariate Taylor series expansion about the mean value of each random variable leads to the law of propagation of uncertainty given by *ANSI*, [3]. The development leads to the evaluation of

$$\sigma^2_{v_p} \; t_A, t_c \; = E \;\; v_p \; t_c \; - \bar{v}_p \; t_A \;^2$$

i.e., the variance of estimating the mean voltage output at t_A from a measurement made at t_c. The voltage is a product of the natural variation in the reference measurements and receiver gain fluctuations.

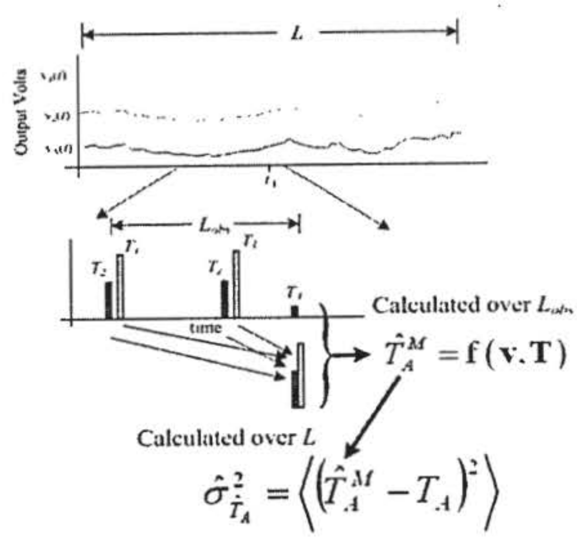


Figure 2: A calibration algorithm **f** *is applied to a data set to form an estimate $\hat{T}_A$. The superscript M denotes the estimate is made from measurements. The algorithm spans an interval L_{obs} and is applied to a data set of length L over which the calculation of the variance in $\hat{T}_A^M$ is made.*

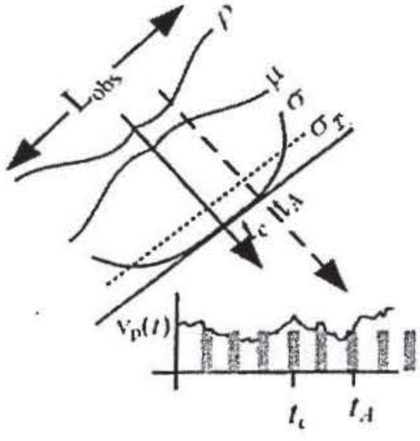


Figure 3: A process is modeled as an array of events. Each event is a random variable with a conditional probability distribution function that is defined to yield the uncertainty of estimating the value of an event at t_A from an event at t_c.

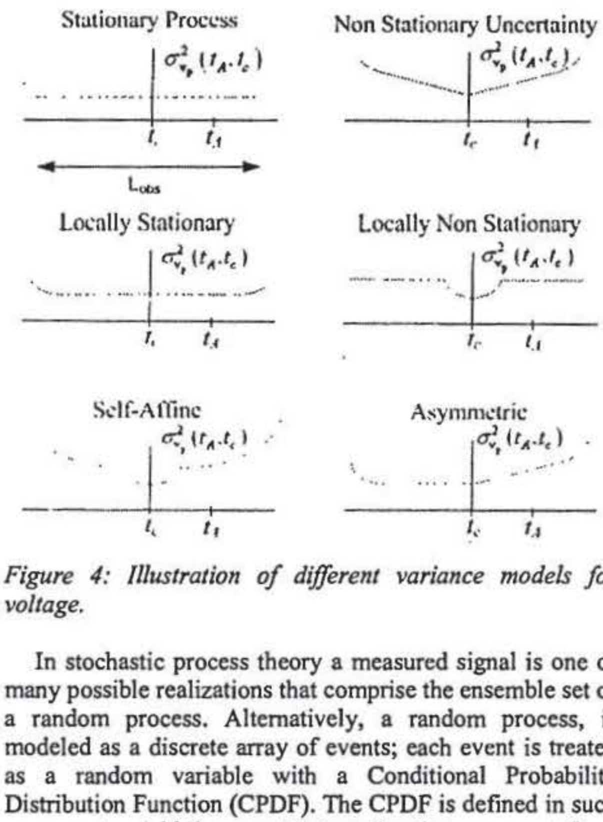


Figure 4: Illustration of different variance models for voltage.

In stochastic process theory a measured signal is one of many possible realizations that comprise the ensemble set of a random process. Alternatively, a random process, is modeled as a discrete array of events; each event is treated as a random variable with a Conditional Probability Distribution Function (CPDF). The CPDF is defined in such a way as to yield the uncertainty estimating one event from another event in the process. Figure 3 illustrates a notional parametric CPDF for an event in an array comprising the process. The CPDF has a stationary component (σ_{T_p}) attributed to the stable noise reference and a non stationary contribution of the receiver fluctuations. Figure 4 illustrates different forms that the variance calculated from the CPDF can take.

4. COMPARISONS WITH DATA

Radiometric data were used with algorithms to investigate the effect of temporal dependencies on the calculation of measurement estimate variance. The data comprise radiometric measurements made at 90 GHz and sequential sampling at 1 Hz rate three blackbody references with temperatures of approximately 80K, 290K, and 328K. The experiment and data set are described in [2]. Herein, a two-pair calibration algorithm is applied to the data set and the variance of the estimator is calculated. The results are shown in Figure 5. The two-pairs of calibration measurements are separated by $2t_I$; the two-pairs are used to estimate the receiver state at t_A; t_A is allowed to vary over +/- 800s; the variance is calculated after applying the algorithm over the ~18000s data set. Three cases are shown for t_I = 6s, 25s and 50s. First note the structure of the variance revealed by the temporal algorithms. If the receiver were stationary, the variance would approximate a straight line; however the structure in the curves reveal detail of the non stationary receiver fluctuations. The variance reaches a minimum when t_A corresponds to the time of calibration and reaches a local maximum between the two calibration pairs. The larger variance for t_I = 6s is attributed to correlation of the receiver fluctuations over this time scale. Note the asymmetry in the uncertainty; the calibration has greater uncertainty projecting into the future than projecting into the past. A five-parameter CPDF fit to the data shows excellent agreement with key features.

5. ENSEMBLE DETECTION AND ANALYSIS

The mathematics and analysis presented above have been developed as a means for comparative analyses of radiometer system architecture and algorithms. The approach attributes radiometer fluctuations to changes in the receiver gain. Although other time-varying factors may contribute to uncertainty in radiometer measurements, the approach shows good promise as a means for detecting, characterizing and modeling fluctuations in radiometer systems.

The mathematical treatment is sufficiently general to apply to the analysis of externally-produced non stationary signals. The approach is a novel form of Noise Assisted Data Analysis (NADA). Figure 6 shows a block diagram of the structural components of EDA. An Ensemble Detector is made from a stable radiometer receiver and stable calibrated noise references. The receiver gain is modulated by an external signal, *g(t)*. The receiver output produces an ensemble set of realizations of *g(t)* to which temporally based algorithms are applied. The calculated variance is compared to analytical parametric

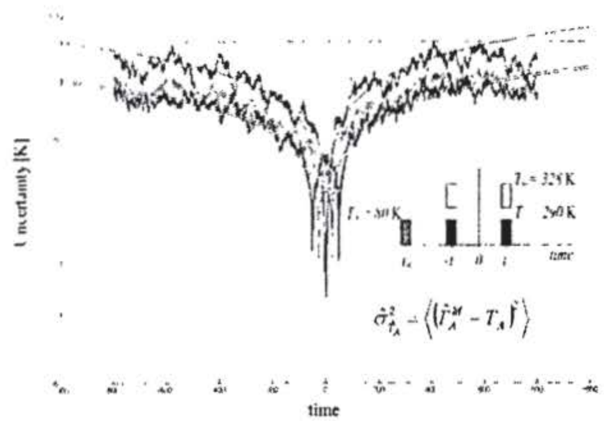
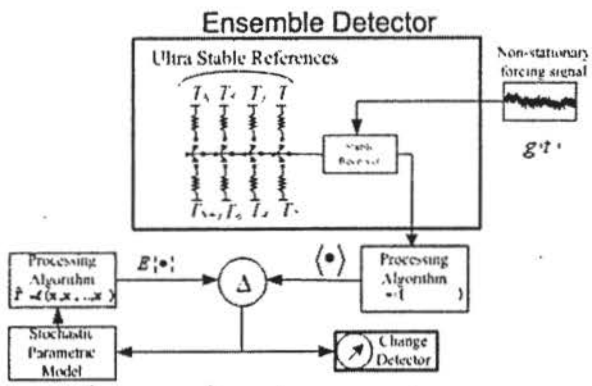


Figure 5: Calculation of the variance in the estimate of $\hat{T}_A$ for variety of temporal algorithm. Calculated variance are fit to a parametric model for the conditional probability distribution function. The non stationary fluctuations in the receiver are very well modeled over a 1600s interval.



$$E\left\{\left(\hat{T}_A - T_A\right)^2\right\} \cong \left\langle\left(\hat{T}_A^M - T_A\right)^2\right\rangle$$

Figure 6: Block diagram illustrating the fundamental components of Ensemble Detection and Analysis.

models of $g(t)$. Comparison between model and measurement permits detection of change and the tuning of parameters that describe the stochastic structure of $g(t)$.

5. COMPEMENT TO EMPERICAL MODE DECOMPOSITION

Over the past decade, Empirical Mode Decomposition (EMD) and the Hilbert-Huang Transform have emerged as leading analysis tools for non stationary data. EMD was developed as an analysis technique for studying the non-stationary and nonlinear propagation of ocean waves [4]. Since its introduction, EMD has grown widely in its application to other fields of science and technologies. However up until recently, the sifting process by which the Intrinsic Mode Functions (IMFs) are derived has had problems associated with an issue known as intermittency. Wu and Huang (2009) [5] report a new Ensemble Empirical Mode Decomposition (EEMD) technique that reduces the effects of intermittency. EEMD is a NADA technique whereby multiple noise series are added to the signal to form an ensemble set on which EMD analysis is performed. The results of the EMD are averaged across the ensemble which results in the reduction of the problem with intermittency. An alternative interpretation of intermittency is that the unique IMFs produced in noisy conditions are different realization of an underlying ensemble set.

EDA is a different yet complementary approach to EEMD. Instead of adding noise to a signal, EDA mixes the signal with calibrated noise. It's the mixing process and the *a priori* statistical relationship of the calibrated noise that permits algorithms to extract temporal information about the signal's underlying stochastic structure. Since EDA is based on the analysis of ensemble data sets and provides the means for comparing measurements with stochastic models, the approach provides a bridge for EMD and stochastic process theory. There exists great potential for EDA when combined with EMD to detect energy transfer between a signal's intrinsic modes. The formulation appears as a macroscopic form of quantum theory, where energy transference is analogous to quantized energy states in atomic structures. Envisioned to emerge from the theoretical framework of EDA is a new Observation Theory.

11. REFERENCES

[1] Racette, P. and R. Lang, "Radiometer design analysis based upon measurement uncertainty," Radio Science, 40, 2005.

[2] P. Racette, "Radiometer Design Analysis Based on Measurement Uncertainty," Dissertation submitted to The George Washington University, May 2005.

[3] ANSI, *American National Standard for Calibration – U.S. Guide to the expression of uncertainty in measurement*, ANSI/NCSL Z540-2-1997, NCSL International, Boulder, CO, 1997.

[4] N. E. Huang, , Z. Shen, S. R. Long, M. C. Wu, H. S. Hsing, Q. Zheng, N-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert Spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society*, 454, pp. 903-995, 1998.

[5] Z. Wu and N. E. Huang, "Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method," Advances in Adaptive Analysis, Vol. 1, No. 1, pp. 1-41, 2009.